Proposed Design Changes to RCRAInfo to Address the WIN/Informed Universe Identification and Waste Activity Monitoring (UID/WAM) Program Area Analysis (PAA) Results

Prepared by:
RCRAInfo UID/WAM Design Team

June 15, 2001

Table of Contents

Background

Now that the migration from RCRIS and BRS to RCRAInfo is complete, incremental re-design, development and implementation of changes to RCRAInfo will begin as each of the WIN/INFORMED Program Area Analyses (PAAs) are completed. The PAAs are the process by which the existing information is assessed and the future information needs of the RCRA program are determined in an effort to make high-quality hazardous waste information available to EPA and States[1], which can use the data to support more effective implementation of the RCRA hazardous waste program. The first completed PAA that recommends changes to RCRAInfo is the Universe Identification/Waste Activity Monitoring (UID/WAM) PAA.

The purpose of this document is to communicate the proposed design changes to RCRAInfo to address the WIN/INFORMED UID/WAM PAA results.

Project Scope

The RCRAInfo UID/WAM Design Team is tasked with implementing the results of the UID/WAM PAA (refer to **Appendix H** for a list of Design Team members). In addition, the Design Team is tasked with developing the means to host a web-based version of the new ***RCRA Subtitle C Site Identification Form*** (***Site ID Form***) and manage processing of the data by States and EPA via the National Exchange Network and the Central Data Exchange (CDX) internet portal. This part of the project is a joint effort between the EPA's Office of Solid Waste (OSW) and the Office of Environmental Information (OEI).

An important goal of the WIN/INFORMED initiative is to identify information that must be collected and made available to all organizations (EPA Headquarters (HQ), EPA Regions and States) to support the implementation of the RCRA program. This information is defined under WIN/INFORMED as *"National"*. Other classifications of information needs are defined under WIN/INFORMED as *"Shared"* and *"Organizational-Specific"*. *National* information is mandatory and *Shared* information is optional, but, both must be nationally defined and have a precise definition that means the same thing to all organizations. *Organizational-Specific* information is defined under WIN/INFORMED as the information needs of any organization, other than those needs determined to be *National* or *Shared.* National systems such as RCRAInfo must accommodate both *National* and *Shared* information needs, while many *Organizational-Specific* needs must be addressed using the organization's own information systems since the national system cannot be designed to meet all needs of all organizations (as we learned with RCRIS). Please refer to the *"WIN/INFORMED Universe Identification and Waste Activity Monitoring Program Area Analysis Final Report"* (January 2000) for precise definitions and additional details regarding *National*, *Shared* and *Organizational-Specific* information needs (http://www.epa.gov/epaoswer/hazwaste/data/win/r00-004.pdf).

---

[1] The term "State" includes American Samoa, the District of Columbia, Guam, the Navajo Nation, Puerto Rico, the Trust Territories, and the Virgin Islands.

The UID/WAM PAA Team used the above-noted definitions of ***National***, ***Shared*** and ***Organizational-Specific*** when developing their recommendations for changes to RCRAInfo. As part of the task to implement the results of the UID/WAM PAA, the RCRAInfo UID/WAM Design Team analyzed the ***"RCRAInfo System Changes"*** documents provided by the PAA Team. These documents provided the PAA Team's recommendations for changes to the existing RCRAInfo database, data entry software, reports and translator file specifications. The PAA Team was not tasked with providing the Design Team precise specifications for system changes, but rather input to assist the Design Team, who is tasked with determining how the recommendations should best be implemented within the system.

Although the UID/WAM PAA Final Report was distributed to the entire WIN/INFORMED community for review and comments, the PAA Team's ***"RCRAInfo System Changes"*** documents were developed specifically for the Design Team and were not distributed to the RCRAInfo User Community for review or comments (a copy of the ***"RCRAInfo System Changes"*** document that details the types of changes the PAA Team recommends is available in the UID/WAM section of the RCRAInfo website (http://www.epa.gov/oswfiles/rcrainfo/). During the Design Team's review of the PAA Team's ***"RCRAInfo System Changes"*** documents, the Design Team became concerned that many of the States and Regions were not aware of the impact of some of the proposed changes to RCRAInfo. The PAA Team's proposed changes consisted of adding new data elements, modifying existing data elements and archiving/removing[2] data elements that the PAA did not identify as ***National*** or ***Shared*** needs. The Design Team feels strongly that some of the data elements identified by the PAA Team to be ***Organizational-Specific,*** and therefore archived/removed from RCRAInfo data entry screens, should still remain available for data entry in RCRAInfo because although a State or Region may not need to "share" this information with another State, Region or HQ, a number of States and Regions "share" the need to track and report on this information for their organization. The data elements slated for removal by the PAA Team that are of concern to the Design Team include the following:

- STATE_DISTRICT
- ACKNOWLEDGE_FLAG
- ACKNOWLEDGE_DATE
- Owner/Operator Address data elements

The Design Team respects the significant work the UID/WAM PAA Team put forth to perform their analyses and develop their recommendations, but, we want the RCRAInfo User Community to understand that Implementers that wish to track the data elements that are archived/removed from RCRAInfo data entry screens will be required to track them in their own organization's information systems, or they must create and maintain their own tables in the RCRAInfo National Database. This is not an easy or straightforward process since it involves significant coordination and interaction between the State/Region, HQ OSW/IMB and EPA's National Technology Services Division (NTSD) in Research Triangle Park, NC to create and maintain the Implementer-developed tables. Please note that HQ OSW/IMB does not have resources to create or maintain these Implementer-developed tables separate from the national RCRAInfo application.

---

[2] **Archive/Remove = data fields no longer available via RCRAInfo data entry screens or translator data loads, but, data is retained and available by querying the RCRAInfo National Database directly using SQL or other reporting tools.**

One of the other goals of the WIN/INFORMED initiative is to reduce burden, and maintaining *Organizational-Specific* data elements in the national system is a cost and resource "burden" to HQ/OSW/IMB, however, the Design Team feels strongly that keeping the existing *Organizational-Specific* data elements that can be **proven to be widely used by Implementers** accessible via RCRAInfo data entry screens is the least burdensome option.

The Design Team is encouraging comments from the RCRAInfo User Community regarding this issue when submitting your comments on the proposed re-design of RCRAInfo. Until a final decision is made on this issue, the re-design of the Handler Module still reflects the *Organizational-Specific* data elements mentioned above that were of specific concern to the Design Team.

RCRAInfo User Community Comments on the Proposed Re-Design

Comments regarding these proposed design changes should be submitted to Dina Villari (villari.dina@epa.gov) no later than **close of business on Tuesday, June 26, 2001**.

Two national conference calls have been scheduled with the RCRAInfo UID/WAM Design Team to answer any questions and receive your feedback on the proposed re-design. The first call will be Tuesday, June 19 from 12:30pm - 2:30pm (ET) and the second call will be Tuesday, June 26 from 12:30pm - 2:00pm (ET). The dial-in number will be the same for both calls, 202-260-1015, conference code 0368#. We welcome all States and Regions to participate in one or both calls, but, there are a limited number of lines available for each call, so we must limit each State and Region to **ONE** line per call.

Please NOTE: The UID/WAM PAA Team is currently preparing a document for the WIN/INFORMED Community that will summarize the PAA Team's recommendations, including an explanation of the rationale behind their decisions (document currently scheduled for release in July). The RCRAInfo Design Team has NOT attempted to provide their rationale in this document, therefore, comments and questions regarding why certain data elements were added, modified, or targeted for archival/removal will be forwarded to the UID/WAM PAA Team for response and will not be included in the Design Team's "Response to User Community Comments" document that will be developed and distributed in late July/early August 2001.

Design Summary: Handler Module

The following summarizes the proposed changes to the RCRAInfo Handler Module. Refer to **Appendix A** for a structure chart outlining the Design Team's proposed changes to the Handler Module and **Appendix B** for a crosswalk of the current structure to the proposed structure.

Site Identification Form

The existing *Notification of Regulated Activity Form* (8700-12), and the site identification portions of the *Hazardous Waste Permit Application, Part A* (8700-28) and the *Hazardous Waste Report Identification and Certification (IC) Form* (8700-13A/B) will be replaced with a new form titled *RCRA Subtitle C Site Identification Form* (refer to **Appendix E** for a copy of the *Site ID Form*). Implementation of the Site ID Form is expected to begin in January 2001 when the *2001 Hazardous Waste Report, Instructions and Forms* have been distributed to the regulated community for use in meeting their 2001 Biennial Reporting requirements. The *Site ID Form* will begin to be used for Notification and Part A reporting requirements as soon as the RCRAInfo software is able to accommodate the data (currently estimated to be March 2002). The remainder of the Part A Permitting Application is currently under development by HQ/OSW's Permits and State Programs Division (PSPD). It is not known at this time what modifications PSPD may make to the Part A Form that will impact the RCRAInfo Permitting Module. The Design Team expects to receive this information from PSPD by early July and will communicate this to the RCRAInfo User Community.

The new *Site ID Form* contains new data fields that require history to be maintained. To implement these recommendations in a way that facilitates the storage of the data, the Design Team proposes saving the entire *Site ID Form* as one "record" or "snapshot", thus maintaining history for everything. Each time a site submits a *Site ID Form*, the information will be entered into the **HHandler Table** as a separate record, using a sequence number to track history. The **HPrevious_Name Table** is no longer necessary and will be archived/removed from RCRAInfo.

A data field was added to the *Site ID Form* to capture the Country for the Site Mailing Address. A new MAIL_COUNTRY field was added to **HHandler Table**.

Because the new *Site ID Form* will also be used for Hazardous Waste Reports, new data fields were added to the Handler Module (Number_of_Employees, Include_In_National_Report and Site_Name) and a link was added from the Handler Module's **HHandler Table** to the **BReport_Cycle Table** in the new Waste Activity Reporting Module. The Include_In_National_Report field in the **HHandler Table** of the Handler Module will be used by Implementers to designate at a **facility level** those facilities that should be included in or excluded from National Biennial Reports.

To track multiple people certifying the accuracy and completeness of their *Site ID Form*, the **HCertification Table** was added to the Handler Module.

The PAA Team changed the name and definition of the existing Owner_Operator_Change_Date field in the **HOwner_Operator Table**. The definition for this field now means that you must fill in the date when the owner or operator *became the current* owner or operator. This is different from the existing definition which requires that you enter a date only when the existing owner or operator is *no longer the current* owner or operator. The name of the new field will be Ownerop_Date_Became_Current . A field identified by the PAA Team as *Shared*, Dunn_Bradstreet_Number , was also added to the **HOwner_Operator Table**.

Contact Information

Following implementation of RCRAInfo, many Users have tried to associate contact information with a source from an activity, rather than the source of the record (e.g., N, A, R) as it was intended. In an effort to reduce this confusion, the Design Team proposes to merge the existing **HContact Table** into the new **HHandler Table**, thus associating contact information with the source of the record as it was originally intended. This change also eliminates the need for an Implementer-defined contact type.

Latitude and Longitude Data

To comply with EPA's locational data standards, additional fields were added to store locational data for all RCRA sites. The existing **HLatitude_Longitude Table** was also merged into the new **HHandler Table**. The **HLatitude_Longitude Table** will be archived/removed from RCRAInfo.

Source Data

The PAA Team recommended consolidating the existing "**E** - **EPA**" and "**S** - **State**" sources into one source called "**I** - **Implementer**". All Handler data entry and translation will be based on the Handler Implementer of Record (IOR) settings. To identify the reason the *Site ID Form* was submitted by the regulated community, the allowed values for Source_Type will be: **'N'** - **Notification**; **'A'** - **Part A**; **'R'** - **Hazardous Waste Report**; **'B'** - **Both Notification and Hazardous Waste Report; 'E'** - **Emergency;** and **'T'** - **Temporary.** The existing **HSource Table** was also merged into the **HHandler Table** and the **HSource Table** will be archived/removed from RCRAInfo.

Activity Data

The PAA Team recommended changing the format of the activity fields to make them consistent and defined all activities as "YES/NO" fields. In addition, new activity fields were added: Importer_Activity, Universal_Waste_Activity, Transfer_Facility, Universal_Waste_Dest_Facility, and Mixed_Waste_Generator (from the Biennial Report). The **HActivity Table** was also merged into the **HHandler Table** and the **HActivity Table** will be archived/removed from RCRAInfo.

Universal Waste Activity will be captured in a separate table called **HUniversal_Waste Table**. This waste type will include a Generated Indicator and/or an Accumulated Indicator and will have it's own lookup table (**LU_Universal_Waste_Activity)** for Implementer-defined wastes.

Generator Status

Every RCRA site will now have *two* generator statuses, one State and one Federal. States will report their status as defined by their own regulatory definitions and they will report the Federal status as defined by the Federal regulations. These two fields will be derived by two new lookup tables (**LU_State_Generator_Activity** and **LU_Federal_Generator_Activity**) that will be linked to the **HHandler Table**.

Conversion from SIC to NAICS Codes

The **HSIC Table** will be replaced with the new **HNAICS Table**. The PAA Team will provide the mapping of the current SIC codes to the NAICS Codes.

Environmental Priority Ranking Table

The **Henv_Priority_Rank Table** will be archived/removed from RCRAInfo, but, will be available for reporting for those few States/Regions that still refer to that data.

Tracking Hazardous Waste Exports

The PAA Team recommends integrating hazardous waste export data from OECA's Hazardous Waste Exports System (HWES) into RCRAInfo on a biennial basis. An additional field Destination_Country will be added to the **HHandler Table** of RCRAInfo to accommodate this information at a later date (2002) when the logistics of how and when this data will integrated into RCRAInfo are determined.

Tracking Hazardous Waste Imports

The PAA Team plans to add a new data element, Country_Of_Origin, to the *Site ID Form* when the Information Collection Request (ICR) is renewed in 2003, but, the data element will not be added to RCRAInfo at this time.

Complete List of Data Elements to be Archived/Removed

Please refer to the Crosswalk in **Appendix B** for a complete list of data elements that will be archived/removed from the existing Handler Module.

Conversion Issues

Conversion Issues will be dealt with at a later date when the design is final and development of the conversion programs is underway.

<u>Design Summary: Waste Activity Reporting Module</u>

The following summarizes the proposed changes to the RCRAInfo Biennial Report Module. Refer to **Appendix C** for a structure chart outlining the Design Team's proposed Waste Activity Reporting Module and **Appendix D** for a Crosswalk of the current structure to the proposed structure.

The design of the new Waste Activity Reporting (WAR) Module was driven by three primary needs:

1.   Adjust the database structure to accommodate the new data requirements based on the ***2001 Hazardous Waste Report, Instructions and Forms (2001 BR Forms)***;

2.   Integrate waste activity data into the new Handler Module structure; and

3.   Integrate the UID/WAM PAA Team's recommendations for improving data quality.

The changes to the data requirements for the ***2001 BR Forms*** include adoption of the new ***RCRA Subtitle C Site Identification Form (Site ID Form)***, which will replace the existing ***Identification and Certification (IC) Form*** beginning with the 2001 Biennial Reporting (BR) cycle. Refer to **Appendix E** for a copy of the new ***Site ID Form***. The new WAR Module data structure integrates into the revised Handler Module structure by storing the ***Site ID Form*** submitted as part of a Biennial Report submission in the **HHandler Table**, and then referencing the **HHandler Table** for the other WAR Module data tables. This effort to integrate the Biennial Report data also includes removing the separate Comments and Descriptions Tables in favor of using various Notes and Descriptions columns used for RCRAInfo data entry.

Changes to existing coding structures for Form Codes, Source Codes, and Management Method Codes will be accomplished by changes to the existing lookup tables and the creation of new lookup tables. Refer to the ***2001 BR Forms*** for complete lists of the new Form Codes, Source Codes, and Management Method Codes (http://www.epa.gov/epaoswer/hazwaste/data/brs01/forms.htm). The Data Elements and Codes used in previous BR cycles will be archived/removed. Please NOTE that Source Code data is now ***National*** and therefore mandatory.

Several data elements previously collected have been deleted (Point of Measure, Offsite Availability Code, SIC Code) from Form GM and will be archived/removed. The NAICS Codes which replace the SIC Codes will be accessible via the new ***Site ID Form***. The PAA Team also deleted Origin Code from Form GM and recommended deletion of the corresponding data element in RCRAInfo. The Design Team is exploring whether the waste **Origin Code** can be accurately derived from the new Source Codes for reporting purposes, and has therefore, not yet removed the **Lu_Origin_Code Table** from the new WAR Module structure. This issue will be discussed in more detail at the June Design Team meeting.

The structure of the WAR Module was also changed to archive/remove derived data elements in the **BWaste_Information Table** that were used in previous Biennial Reporting cycles, but will no longer be used (e.g., Waste_Water_Calc and Waste_Regulatory_Status_Calc). Please refer to the Crosswalk in **Appendix D** for a complete list of **BWaste_Information Table** data elements that will be archived/removed.

To accommodate the PAA Team's recommendations for improving data quality, erroneous data will no longer be accepted for optional data elements (e.g., Form Code on Form GM). All optional data elements must either meet the nationally defined code lists, or not be provided to EPA for inclusion in the RCRAInfo National Database. This requirement will be implemented with a combination of lookup tables and edit checks in the BR data load software in RCRAInfo.

The Include_In_National_Report field in the **BGM_Basic_Information and BWR _Basic_Information Table** of the WAR Module will be used by Implementers to designate, **at the waste stream level**, those wastes that should be included in or excluded from National Biennial Reports.

Complete List of Data Elements to be Archived/Removed

Please refer to the Crosswalk in **Appendix D** for a complete list of data elements that will be archived/removed from the existing Biennial Report Module.

Conversion Issues

Conversion Issues will be dealt with at a later date when the design is final and development of the conversion programs is underway.

Please NOTE: The UID/WAM PAA Team will provide a complete summary of all changes to the 2001 BR Forms in the document they are preparing to summarize their recommendations.

Design Summary: Central Data Exchange Site Notification Project

The Site Notification Project (SNP) is a cooperative effort between EPA's Office of Solid Waste and the Office of Environmental Information. The purpose of the SNP is to successfuly submit a electronically created Site ID Form to the EPA through the National Exchange Network and the CDX Internet portal.

OSW has taken the lead in developing and maintaining the software that will allow the regulated community to submit their Site ID Forms via the Internet. The SNP software will also consist of an "approval" module that will allow participating States and Regions to accept and reject the submissions.

CDX will be responsible for maintaining the portal (both hardware and software components) and allowing the regulated community access to the application. Currently, the CDX internet portal resides on a third party platform outside the EPA firewall.

RCRAInfo will pilot this new functionality with a group of States and Regions beginning in January 2002, when the new Site ID Form is adopted.

Refer to **Appendix F** for a flow diagram of the planned RCRAInfo/CDX Site Notification process.

Summary of Functionality

Essentially, the SNP is divided into two steps.

1. **Data Entry**

   The regulated community will log onto the CDX portal and select a hyperlink to go to the Site ID Form. They will enter their data into the form. When a submit button is clicked, client-side (i.e., regulated community) data entry validation will be performed to ensure the quality of the data. If the client-side validation fails, then messages are displayed informing the user of the errors. Once client-side quality checks are met, the form is sent to the server where more validation is performed. Again, if validation checks are violated, the messages are returned to the user.

   After passing both sets of edit validation checks, the data is entered into staging tables prior to being accepted into RCRAInfo. A confirmation page is presented to the user. The user must print and sign the confirmation page and send it to the regulating agency.

2. **Data Acceptance**

   After the data is accepted into the staging tables and the confirmation form received, participating States will be allowed to verify that the electronically submitted form is valid and should be accepted into RCRAInfo. If the participating State determines the data to be invalid, then the record is rejected and the regulated community user is notified. Otherwise, the data is uploaded into RCRAInfo.

Impact on RCRAInfo

The expected impact on the existing RCRAInfo application is detailed below.

The Data Entry portion of the SNP will reside on CDX hardware. Preliminary research indicates that the following changes to RCRAInfo will be required:

- the creation of staging tables within RCRAInfo data structure;

- the creation of a "tunnel" between the CDX portal and the RCRAInfo database; and

- the creation of a reporting mechanism to ensure the regulated community can determine where in the pipeline their submissions are located.

The Data Acceptance portion of the SNP will also require changes to RCRAInfo, most notably:

- the addition of another "module" to the RCRAInfo application, which will also impact some of the existing menus, and allow only authorized RCRAInfo users the ability to translate the regulated community's submission into the Handler Module structure;

- modifications to the RCRAInfo User Maintenance Table structure to indicate appropriate user permissions; and

- the creation of reports to reflect various aspects of the submission process (e.g., statistical reports outlining how many submissions were accepted versus rejected or analytical reports outlining which facilities use this process).

We also anticipate having to address Security issues as part of both the Data Entry and Data Acceptance portions of the SNP.

UID/WAM Reporting Views

The PAA Team released a draft document, "Access System Prototype", to the Design Team which proposes approaches for four categories of information requests:

- Site Profile
- Targeted Sites
- Universe Counts
- Waste Quantity

The PAA Team will compile the final report recommendations and provide them to the Design Team for consideration and development later this summer, but in the interim, the Design Team is proposing two approaches to reporting:

1. Development of a Handler/Waste Activity Reporting view or table; and
2. Counts in a matrix format with a drill down capability.

1. **Reporting View or Table**

    Although the newly proposed Handler Module has a correct relational database structure, it is too complicated for convenient reporting. Often the best structure to store data for data entry does not allow for efficient reporting. Therefore, a more simplified reporting structure must be developed to simplify reporting and maximize query speed.

    There are several precedents for this approach in RCRAInfo. A reporting table is in place that simplifies reporting for Compliance, Monitoring and Enforcement data: CMECOMP and Biennial Reporting data: BWaste_Information. The CMECOMP Table, which is updated concurrently with the base tables as data is added, updated or deleted, combines all the CM&E data into one table, rather than the three main CM&E tables plus various linking and lookup tables. CMECOMP also contains a lot of Handler Module data (e.g., handler name, county, location and mail addresses, non-notifier indicator, state district) for further reporting convenience. In addition to simplifying query logic, CMECOMP provides faster performance in some cases (depending on which fields are indexed) than reporting from the base tables because the need for table-to-table joins is eliminated. Joins tend to slow queries.

    A View is simply stored SQL logic. RCRAInfo has a view, V_HUNIVERSE, that relieves the report developer from having to write SQL to retrieve the various universes that any one handler belongs in. The Huniverse Table has one record for each universe the handler belongs in, and may have multiple records for any one Handler ID. The select logic of V_HUNIVERSE in effect retrieves the multiple universe records for a handler and puts them into one record.

For a Handler Module reporting structure, we have not yet determined whether a reporting table is necessary or if a view is sufficient, because a view is just stored SQL logic, it does not necessarily result in faster queries. Although we have not yet designed the new reporting structure, a few observations can be made about what it will contain. It will contain only current information, not historical records. Although having historical data is a very valuable resource, the Design Team believes a reporting view/table version of the Handler data that contains only current data will satisfy more than 90% of the reporting needs of the RCRAInfo User Community.

Therefore, the reporting structure will contain current data for:

- handler name and handler ID
- owner and operator
- contact
- all fields in **HHandler Table**, including addresses, lat/long, etc.

Only the most recent records in **HHandler Table** where the activity_location matches the first two characters of the Handler ID number will be included.

Also, all of the calculated national universes and GPRA universes ("baseline" universes) will be incorporated into the structure. Many-to-one Handler data, such as waste codes, process codes and other permits will not be included in the structure. (An exception is NAICS codes; the structure will allow up to four NAICS codes.)

2. **Matrix/Drill Down Reporting**

Because of the lack of resources, at both the HQ and the Implementer level, report development has been slow in RCRAInfo. The most serious deficiency is the lack of reports for managers. Managers are usually interested in counts, but sometimes want to see the details behind the counts. Moving from counts down to the details behind the counts is often referred to as "drill down". This capability has not been used in RCRAInfo reports yet, but is possible with existing reporting tools, Oracle Reports plus Adobe Acrobat files. An Adobe Acrobat file can contain live links that will start another report using the parameters implied by the contents of the matrix cell.

A simple example is contained in **Appendix G**. The matrix report is shown resulting from the user's selection of universe counts for Region 7. The user then decides he wants to see the handlers corresponding to the SNC count of 17 for Nebraska. He clicks on that cell and that action submits a handler universe list report, seen in the second "page" of **Appendix G**. (Please note that these reports are just a rough draft, but do contain real RCRAInfo production data.)

The Design Team believes there will be many potential uses for the "drill down" approach in every RCRAInfo Module, limited only by the availability of detail reports to "drill down" to. As part of implementing the UID/WAM PAA results, universes count reports similar to the example in **Appendix G** will be developed.